AD-A090 186 FLORIDA STATE UNIV TALLAHASSEE DEPT OF STATISTICS F/G 12/1
STOCHASTIC MODELS FOR THE RANDOM LOCATION OF INDIVIDUALS IN A H--ETC(U)
JUN 80 N A LANGBERG, R JOHNSON, R A BRADLEY N00014-76-C-0608
UNCLASSIFIED FSU-STATISTICS-M455R AFOSR-TR-80-0930 NL

1 OF 1
AD
A090186

END
DATE FILMED
11-80
DTIC

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered) UNCLASSIFIED

REPORT DOCUMENTATION PAGE

READ INSTRUCTIONS BEFORE COMPLETING FORM

1. REPORT NUMBER: AFOSR-TR-80-0930
2. GOVT ACCESSION NO.: AD-A090186
3. RECIPIENT'S CATALOG NUMBER
4. TITLE (and Subtitle): STOCHASTIC MODELS FOR THE RANDOM LOCATION OF INDIVIDUALS IN A HABITAT.
5. TYPE OF REPORT & PERIOD COVERED: Interim rept.
6. PERFORMING ORG. REPORT NUMBER
7. AUTHOR(s): Naftali A. Langberg, Roger Johnson, Ralph A. Bradley
8. CONTRACT OR GRANT NUMBER(s): N00014-76-C-0608, N00014-80-C-0093, AFOSR 78-3678
9. PERFORMING ORGANIZATION NAME AND ADDRESS: Florida State University, Department of Statistics, Tallahassee, FL 32306
10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS: 61102F 2304/A5
11. CONTROLLING OFFICE NAME AND ADDRESS: Air Force Office of Scientific Research/NM, Bolling AFB, Washington, DC 20332
12. REPORT DATE: June 1980
13. NUMBER OF PAGES: Ten
14. MONITORING AGENCY NAME & ADDRESS (if different from Controlling Office): 12
15. SECURITY CLASS. (of this report): UNCLASSIFIED
15a. DECLASSIFICATION DOWNGRADING SCHEDULE

LEVEL

16. DISTRIBUTION STATEMENT (of this Report)

Approve for public release; distribution unlimited.

TR-1 9NR

17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report)

DTIC ELECTE OCT 9 1980 C

18. SUPPLEMENTARY NOTES

19. KEY WORDS (Continue on reverse side if necessary and identify by block number)

stationarity, independent increments, compound Poisson, binomial, and negative binomial distributions

20. ABSTRACT (Continue on reverse side if necessary and identify by block number)

Various probabilistic models for the random locations of individuals in a region are considered. Models for both infinite and finite regions are developed. In the literature, certain marginal distributions are used to characterize the numbers of individuals in subregions. Relationships between these marginal distributions and our probabilistic models are discussed.

DD FORM 1473 1 JAN 73 EDITION OF 1 NOV 65 IS OBSOLETE

UNCLASSIFIED 400277

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

AFOSR-TR- 80-0930

STOCHASTIC MODELS FOR THE RANDOM LOCATION OF INDIVIDUALS IN A HABITAT

by

Naftali A. Langberg[1], Roger Johnson, and Ralph A. Bradley[2]

FSU Statistics Report M455R
AFOSR Technical Report No. 78-110
ONR Technical Report No. 132R

June, 1980
The Florida State University
Department of Statistics
Tallahassee, Florida 32306

[1]Research supported by the Air Force Office of Scientific Research, AFSC, USAF, under Grant AFOSR-78-3678.

[2]Research supported in part by the Army, Navy and Air Force under Office of Naval Research Contract No's. N00014-76-C-0608 and N00014-80-C-0093, and in part under Air Force Office of Scientific Research Contract No. AFOSR-78-3678.

Stochastic Models for the Random Location of Individuals in a Habitat*

by

Naftali A. Langberg, Roger B. Johnson, and Ralph A. Bradley

Department of Statistics
Florida State University
Tallahassee, Florida 32306, U.S.A.

Abstract: Various probabilistic models for the random locations of individuals in a region are considered. Models for both infinite and finite regions are developed. In the literature, certain marginal distributions are used to characterize the numbers of individuals in subregions. Relationships between these marginal distributions and our probabilistic models are discussed.

Key Words: stationarity, independent increments, compound Poisson, binomial, and negative binomial distributions.

*Research supported in part by the Army, Navy and Air Force under Office of Naval Research Contract No's. N00014-76-C-0608 and N00014-80-C-0093, and in part under Air Force Office of Scientific Research Contract No. AFOSR-78-3678A. Reproduction in whole or in part is permitted for any purpose of the United States Government.

1. Introduction and Summary.

Trees in a forest, stars in a galaxy and bacteria on a petri dish are a few examples wherein there is interest in the spatial distributions of individuals over regions. In pursuit of better understanding of these distributions, attempts have been made to use probabilistic and statistical tools [Eberhardt (1967), Pielou (1969), Pollard (1971), Patil and Stiteler (1974), Cox (1976), Cox and Lewis (1976)]. Very little effort has been made to develop coherent statistical structures for the locations of individuals in a region. The main purpose of this paper is to propose such structures in order to avoid inconsistencies that can be found in the literature, to provide a sound probabilistic basis for a variety of inference procedures that are used, and to permit the development of better and consistent statistical methodology in the area.

We show that, under certain conditions, the collection of random variables representing the numbers of individuals in subsets of a region can be regarded as a stochastic process. The models presented have a stationarity property. In Section 2 we present simple random models which have, in addition to stationarity, a property of independent increments. We prove that simple models lead to a compound Poisson process. In a number of the references cited above, Poisson and negative binomial distributions have been used to describe the numbers of individuals in subsets of a region. Our formulation shows that these two distributions arise as marginal distributions in special cases of simple random models.

In Section 3 we introduce models that are stationary, but do not necessarily have the property of independent increments. We show that a negative binomial distribution proposed by Eberhardt (1967) to characterize the random behavior of the numbers of individuals in subsets of a region is a marginal

distribution resulting from two different stochastic models that do not have independent increments. Section 4 is devoted to the development of several different stochastic models when the region under consideration has finite Lebesgue measure. It is shown that the binomial distribution suggested by Eberhardt (1967) can be derived from all of these models.

2. Simple Random Models.

Some basic notation is needed to define spatial stochastic models. Throughout the paper, R will denote a Euclidean space, λ will be the Lebesgue measure on R, X(A) will be the number of individuals in a subset $A \subseteq R$, and Σ will represent the set of all Borel subsets of R with finite Lebesgue measures.

Definition 2.1. Individuals are said to be located in R in a simple random way if the following conditions hold:

(i) Stationarity: For every set $A \in \Sigma$, the distribution of the random variable X(A) depends only on $\lambda(A)$.

(ii) Independent increments: If $A_1, \ldots, A_m$ $(m > 1)$ represent any m disjoint sets in Σ, the random variables, $X(A_1), \ldots, X(A_m)$, are independent.

The objective of this section is to show that, under conditions (i) and (ii), the collection, $\{X(A),\ A \in \Sigma\}$, is a stochastic process determined by a positive number and a sequence of numbers in [0, 1] summing to unity.

For any two sets A_1, A_2 in Σ and nonnegative integers k_1, k_2, we have clearly

$$P[X(A_1) = k_1,\ X(A_2) = k_2] = \sum_{\substack{j+r=k_1 \\ j+\ell=k_2}} P[X(A_1 \cap A_2) = j,\ X(A_1 \cap \bar{A}_2) = r,\ X(\bar{A}_1 \cap A_2) = \ell] \tag{2.1}$$

Now let q be an arbitrary point in R and $Y_q(t)$, the number of individuals in a sphere centered at q having Lebesgue measure $t \geq 0$, $Y_q(0) \equiv 0$. Then, using (i), (ii), we can write

$$P[X(A_1) = k_1,\ X(A_2) = k_2] = \sum_{\substack{j+r=k_1 \\ j+\ell=k_2}} P[Y_q(t_1) = j]P[Y_q(t_2) = r]P[Y_q(t_3) = \ell], \quad (2.2)$$

where t_1, t_2 and t_3 are $\lambda(A_1 \cap A_2)$, $\lambda(A_1 \cap \bar{A}_2)$ and $\lambda(\bar{A}_1 \cap A_2)$ respectively. Without difficulty, equations (2.1) and (2.2) can be extended to any finite number of sets in Σ. The general result is stated in the following theorem.

Theorem 2.1. For individuals located in R in a simple random way, the joint distribution of any finite collection of random variables from $\{X(A),\ A \in \Sigma\}$ is determined uniquely by a joint distribution of a finite collection of random variables from $\{Y_q(t),\ t \geq 0\}$.

The proof is omitted. It remains to show that, under (i) and (ii), the collection $\{Y_q(t),\ t \geq 0\}$ is a stochastic process determined by a positive number and a probability sequence.

Conditions (i) and (ii) respectively insure that the process $Y_q(t)$ has stationary and independent increments. That is, for real numbers $0 \leq t_1 \leq t_2 \leq \ldots \leq t_m$ $(m \geq 2)$, the random variables $Y_q(t_1)$, $Y_q(t_2) - Y_q(t_1)$, ..., $Y_q(t_m) - Y_q(t_{m-1})$ are independent and, moreover, $Y_q(t_\ell) - Y_q(t_{\ell-1})$ has a distribution depending only on $t_\ell - t_{\ell-1}$, $\ell = 1, \ldots, m$ $(t_0 = 0)$. Upon application of a well known characterization of such processes, it follows [see, for example, Khintchine (1960), p. 36] that $Y_q(t)$ is the compound Poisson process represented as

$$Y_q(t) = Z_1 + Z_2 + \ldots + Z_{N(t)},$$

where $\{N(t),\ t \geq 0\}$ is a simple Poisson process and $\{Z_i,\ i \geq 1\}$ is a sequence of i.i.d. random variables on the positive integers, independent of $N(t)$. If $Z_i \equiv 1$, then $Y_q(t) \equiv N(t)$, whereas if Z_i has a logarithmic distribution, $Y_q(t)$ has a negative binomial distribution, as is well known. These two marginal distributions of $X(A)$ have been frequently used in the literature.

3. A Generalization of the Simple Random Models.

In this section we present two methods of developing a general stationary stochastic structure for the random quantities $\{X(A)\}$. This is done in two steps. Firstly, for $m \geq 1$, we present the joint distribution of $X(A_1), \ldots, X(A_m)$, where $A_1, \ldots, A_m$ are any m disjoint sets of Σ. We then show that this set of finite dimensional distributions determines uniquely the stochastic structure of $\{X(A)\}$.

For this purpose we introduce some notation. Let $\{M(t),\ t \geq 0\}$ be a stochastic process on the nonnegative integers, stochastically increasing in t, that is, $M(t) \leq M(s)$ for $0 \leq t \leq s$, and $\{W_i,\ i \geq 1\}$, a sequence of random variables with positive integer values, independent of $\{M(t)\}$. We assume that $\{M(t),\ t \geq 0\}$ and $\{W_i,\ i \geq 1\}$ depend on the parameters θ_1 and θ_2 respectively, ranging in parameter spaces $\mathcal{P}_1$ and $\mathcal{P}_2$, where $\mathcal{P}_2$ is a collection of infinite sequences, $\mathcal{P}_1$ and $\mathcal{P}_2$ being probability spaces with F and G the respective probability measures.

Definition 3.1. Let $A_1, \ldots, A_m$ be disjoint sets in Σ with respective Lebesgue measures $t_1, \ldots, t_m$, and let $k_1, \ldots, k_m$ be nonnegative integers. We define the joint distribution of $\{X(A_i),\ i = 1, \ldots, m\}$ as

$$P[X(A_i) = k_i,\ i = 1, \ldots, m] = \int_{\mathcal{P}_1 \times \mathcal{P}_2}\int \prod_{i=1}^{m} P\Big[\sum_{j=1}^{M(t_i)} W_j = k_i\Big]dF(\theta_1)dG(\theta_2). \tag{3.1}$$

Definition 3.2. With the same notation as above, we define

$$\int_{P_1 \times P_2}\int P[\sum_{j=r_{i-1}+1}^{r_i} W_j = k_i,\ i = 1, \ldots, m] dF(\theta_1) dG(\theta_2), \tag{3.2}$$

where $r_i = \sum_{q=1}^{i} M(t_q)$, $i = 1, \ldots, m$, $r_0 = 0$.

The consistency of the set of joint distributions of finite collections of random variables from $\{X(A), A \in \Sigma\}$, generated by disjoint sets and defined by (3.1) or (3.2), follows from the stochastic structure imposed on $\{M(t), t \geq 0, W_i, i \geq 1\}$. The stationarity condition (i) is clearly satisfied, since definitions (3.1) and (3.2) vary only with the Lebesgue measures of the respective sets. The generalization of equation (2.1) to any finite number of sets in Σ, not necessarily disjoint, provides a way to extend definitions (3.1) and (3.2) to any such sets. These extensions preserve the required consistency and stationarity of $\{X(A), A \in \Sigma\}$. We have proved the following:

Theorem 3.1. If the joint distributions of every finite number of random variables in $\{X(A), A \in \Sigma\}$ are given by the extensions of (3.1) or (3.2), then $\{X(A)\}, A \in \Sigma\}$ is a stationary stochastic process.

It is important to note that (3.1) and (3.2) yield the same marginal distributions for $X(A)$, $A \in \Sigma$, but they define different processes. Statistical methodologies based on the two models may be quite different and this has been ignored in the literature.

In Section 2, we proved that simple random models depend on a positive number, say μ, and a probabilistic sequence, say $\{P_i, i \geq 1\}$. Now we show that those models are particular cases of the stochastic processes presented by (3.2). To do so, let $\{M(t), t \geq 0\}$ be a Poisson process with parameter μ, and $\{W_i, i \geq 1\}$ be an i.i.d. sequence of random variables given by $P(W_1 = i) = P_i$ $(i \geq 1)$. In addition we take P_1 and P_2 to be sets containing only μ and the sequence $\{P_i\}$ respectively. Now the extension of (3.2) reduces directly to a simple random model.

Let us assume that the sequence of discrete random variables $\{W_i\}$ is degenerate at 1, P_1 is $[0, \infty)$, P_2 is only the sequence $(1, 0, 0, \ldots)$, $\{M(t)\}$ is Poisson, and F is given by

$$F(x) = \begin{cases} \frac{\beta^\alpha}{\Gamma(\alpha)} \int_0^x e^{-\beta y} y^{\alpha-1} dy & x > 0 \\ 0 & x \le 0 \end{cases} \quad \alpha, \beta > 0. \tag{3.3}$$

Then the marginal distribution of $X(A)$, $A \in \Sigma$, computed by (3.1) or (3.2), is negative binomial with $a = \alpha$ and $b = \frac{\alpha}{\beta} \lambda(A)$. This marginal distribution for $X(A)$ was assumed by Eberhardt (1967) and by Patil and Stiteler (1974).

4. Models for Finite Regions.

In the previous two sections, we presented stochastic models when individuals were randomly located in an infinite region. In this section, five different models are developed for situations wherein individuals are located in a Borel set $R_f \subseteq R$ of finite Lebesgue measure.

Let Σ_f be the set of all Borel sets contained in R_f and $X_f(A)$, the number of individuals in set A, $A \in \Sigma_f$. Our objective is to develop a stochastic structure for the collection $\{X_f(A), A \in \Sigma_f\}$. One way to achieve this objective is to select a stochastic process developed in Section 3 for the infinite region R, say $\{X(A), A \in \Sigma\}$, and condition it by an event related to the random variable $X(R_f)$. Specifically, let $A_1, \ldots, A_m$ be sets in Σ_f and let $k_1, \ldots, k_m$ be nonnegative integers. We define the desired probability as

$$P[X_f(A_i) = k_i,\ i = 1, \ldots, m] = P[X(A_i) = k_i,\ i = 1, \ldots, m \mid X(R_f)]. \tag{4.1}$$

The consistency and stationarity of the process $\{X_f(A), A \in \Sigma_f\}$ defined in (4.1) are self-evident.

For the four remaining alternative ways of developing a stochastic structure for $\{X_f(A)\}$, we refer to the process $\{M(t)\}$, the sequence $\{W_i\}$, the sets P_1 and P_2, and the probability measures F and G, which were introduced in Section 3. We define first the joint distribution of $X_f(A_1), \ldots, X_f(A_m)$, where $A_1, \ldots, A_m$ are disjoint sets in Σ_f and $m \geq 1$, then extend these definitions to any finite collection of sets in Σ_f. Since the extension technique has been used twice, it is omitted from our present discussion. Now let $A_1, \ldots, A_m$ be disjoint subsets of R_f, $\lambda(R_f) = t_0$, $\lambda(A_i) = t_i$, $i = 1, \ldots, m$, and let $k_1, \ldots, k_m$ be nonnegative integers. We define the desired probability in the following four ways:

$$P[X_f(A_i) = k_i,\ i = 1, \ldots, m] = \int\int_{P_1 \times P_2} \prod_{i=1}^{m} P\Big(\sum_{j=1}^{M(t_i)} W_j = k_i \Big| \sum_{j=1}^{M(t_0)} W_j\Big) dF(\theta_1) dG(\theta_2). \quad (4.2)$$

$$P[X_f(A_i) = k_i,\ i = 1, \ldots, m] = \int\int_{P_1 \times P_2} P\Big(\sum_{j=r_{i-1}+1}^{r_i} W_j = k_i,\ i = 1, \ldots, m \Big| \sum_{j=1}^{M(t_0)} W_j\Big) dF(\theta_1) dG(\theta_2). \quad (4.3)$$

$$P[X_f(A_i) = k_i,\ i = 1, \ldots, m] = \int\int_{P_1 \times P_2} \prod_{i=1}^{m} P\Big(\sum_{j=1}^{M(t_i)} W_j = k_i \Big| M(t_0)\Big) dF(\theta_1) dG(\theta_2). \quad (4.4)$$

$$P[X_f(A_i) = k_i,\ i = 1, \ldots, m] = \int\int_{P_1 \times P_2} P\Big(\sum_{j=r_{i-1}+1}^{r_i} W_j = k_i,\ i = 1, \ldots, m \Big| M(t_0)\Big) dF(\theta_1) dG(\theta_2). \quad (4.5)$$

where $r_0, \ldots, r_m$ are as in definition 3.2.

The stochastic processes determined by (4.2)-(4.5) have been constructed to be both consistent and stationary. Definitions (4.2) and (4.3) or (4.4) and (4.5), yield identical marginal distributions for $X_f(A)$, $A \in \Sigma_f$, but define different processes.

If we let $M(t)$ be a Poisson process, $\{W_i\}$, a sequence of random variables degenerate at 1, P_1 and P_2 singleton sets, then the marginal distribution of $X_f(A)$, $A \in \Sigma_f$, according to each of the definitions (4.1) - (4.5) is binomial with $X(R_f)$ corresponding to the number of Bernoulli trials and $\lambda(A)/\lambda(R_f)$ corresponding to the probability of success in a single trial.

5. Concluding Remarks.

The selection of models for stochastic processes for particular applications and the development of pertinent statistical methodologies have not been addressed in this paper. We have demonstrated that models proposed may be used to yield marginal distributions assumed in the literature; they may also be used to avoid unwarranted assumptions and inconsistencies that arise. We propose in subsequent work to use the general models of this paper to devise improved statistical methodologies for problems involving the location of individuals in a habitat.

REFERENCES

[1] Cox, T. F. (1976). The robust estimation of the density of a forest stand using a new conditioned distance method. Biometrika, 63, 493-499.

[2] Cox, T. F. and T. Lewis (1976). A conditioned distance ratio method for analyzing spatial patterns. Biometrika, 63, 483-491.

[3] Eberhardt, L. L. (1967). Some developments in distance sampling. Biometrics, 23, 207-216.

[4] Khintchine, A. Y. (1960). Mathematical Methods in the Theory of Queueing, Hafner Publishing Company, New York.

[5] Patil, G. P. and W. M. Stiteler (1974). Concepts of aggregation and their quantification. A critical review with some new results and applications. Res. Popul. Ecol., 15, 238-253.

[6] Pielou, E. C. (1969). An Introduction to Mathematical Ecology, John Wiley and Sons, New York.

[7] Pollard, J. H. (1971). On distance estimators of density in randomly distributed forests. Biometrics, 27, 991-1002.

MED

8